Foundations for learning and adaptation in a multi degree of freedom unmanned ground vehicle

Michael R. Blackburn[1,a] and Richard Bailey[b]
[a]SSC San Diego, 53560 Hull Street, San Diego, CA, USA 92152
[b]ACEi, 25133 Avenue Tibbits, Unit A, Valencia, CA, USA 91355

ABSTRACT

The real-time coordination and control of a many motion degrees of freedom (dof) unmanned ground vehicle under dynamic conditions in a complex environment is nearly impossible for a human operator to accomplish. Needed are adaptive on-board mechanisms to quickly complete sensor-effector loops to maintain balance and leverage. This paper contains a description of our approach to the control problem for a small unmanned ground vehicle with six dof in the three spatial dimensions. Vehicle control is based upon five fixed action patterns that exercise all of the motion dof of which the vehicle is capable, and five basic reactive behaviors that protect the vehicle during operation. The reactive behaviors demonstrate short-term adaptations. The learning processes for long-term adaptations of the vehicle control functions that we are implementing are composed of classical and operant conditionings of novel responses to information available from distance sensors (vision and audition) built upon the pre-defined fixed action patterns. The fixed action patterns are in turn modulated by the pre-defined low-level reactive behaviors that, as unconditioned responses, continuously serve to maintain the viability of the robot during the activations of the fixed action patterns, and of the higher-order (conditioned) behaviors. The sensors of the internal environment that govern the low-level reactive behaviors also serve as the criteria for operant conditioning, and satisfy the requirement for basic behavioral motivation.

Keywords: UGV, Mobility, Learning, Autonomy, Motivation

1. INTRODUCTION

Most existing unmanned vehicles are controlled by teleoperation. The human operator, usually through a joystick and radio link, directs a robot's single degree of freedom, or multiple degrees of freedom sequentially, to execute some maneuver in a complex obstacle-rich environment. Humans require intensive training, often taking years, to manage the coordination of more than two degrees of freedom (for example – in playing the piano). Because of this human cognitive/performance limitation, the use of small unmanned ground vehicles with sufficient degrees of motion freedom for operation in tactical situations involving obstacle dense natural terrain will likely not be possible without competent and adaptive control processes resident on the vehicle[2]. It is to this requirement that the present effort is dedicated.

The present approach builds upon an idea that is at least several hundred millions of years old. This idea is that agent intelligence must develop from processes that promote the survival of the agent. We took this idea and first built the robot shown in Figure 1, adding elements we anticipated to be necessary (but yet insufficient) to develop an intelligent adaptive controller. The elements are multiple degrees of motion freedom, and sensors of critical events in the internal and external environments. Needed to complete the elements, and of which we report herein, are hard-wired fixed action patterns, semi-modifiable basic reactive patterns, and the mechanisms by which our robot agent will be able to acquire mobility and survival skills. Our control architecture containing these elements should permit the acquisition of novel behavioral patterns by the robot to improve its adaptation to its environment.

Our approach should result in a very different kind of an artificial agent. Because our aim with this work is to lay the essential foundation for all higher-level intelligent processes that emulate the biological, when successful we will be well-prepared to explore methods for decision making and tactical behaviors in the agent that are required for collaboration with other unmanned systems, and with humans.

[1] mike@spawar.navy.mil, phone (619) 553-1904; fax (619) 553-6188
[2] In addition, radio-frequency communication limitations will have negative consequences for remote control of unmanned vehicles in complex scenarios.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **APR 2004**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE
Foundations for Learning and Adaptation In A Multi Degree of Freedom Unmanned Ground Vehicle

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
SSC San Diego 53560 Hull Street San Diego, CA 92152

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**10**	

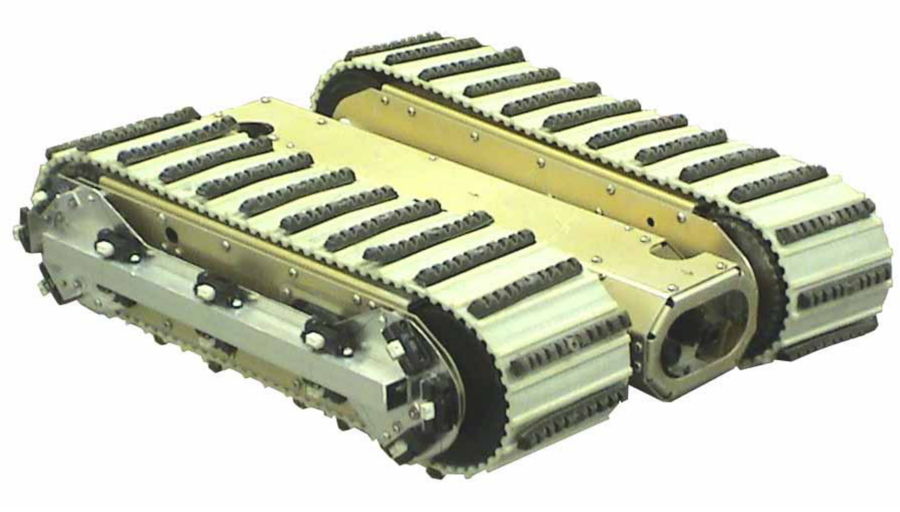

Figure 1. Outer appearance of the Novel UGV

2. ARCHITECTURE

To achieve our objectives, our control architecture must differ in several significant ways from the principal approach taken today in mobile robotics. First, we view our robot as an independent agent, and have attempted to endow it with the necessary capabilities to promote its own welfare. Second, we view our own role in the operation of the robot more as director and collaborator, than as user and operator. In essence, the robot should have eventual control over its own on-off switch, and perform work to further its own survival.

The control processes (algorithms) for our robot must execute within the constraints imposed upon it by our mechanical design, sensors, electronics, and a few behavioral tendencies that we as designers have the privilege to define. All of these things we provide to the robot in assembly, and are analogous to the ontological implementation of a genetic code. If we are foresightful, we will have endowed our robot with the necessary equipment to accomplish survival-based objectives in the envisioned operational environment.

The Novel UGV, shown in Figure 1, is composed of three principal segments, a central core, and two tracked pods. All three segments contain electrical power, power transmission mechanisms, sensors for both the internal and external environments, radios for inter-pod communication, and electronics for local processing. The core manages radio communications with the operator control unit (OCU). The pods are each connected to the central core by a single L-shaped axle, about which the pods can rotate. These two axles are mounted at either end of the core, and laterally near the end of each pod. The axles, with pods attached, can rotate 180 degrees about the ends of the core.

The NUGV is symmetrical on all major axes, so that if the image in Figure 1 was rotated 180 degrees in any direction, it would appear the same. Sensors for the external environment (video cameras, and IR range sensors) are located on both ends of the core faceplates, and (sans video cameras) on the outboard sides of the two pods.

The physical architecture of our robot permits it to assume several different conformations. A sample of the different conformations that are possible with the Novel UGV's six degrees of freedom is given in Figure 2. The variable conformation of the vehicle permits a large diversity of behavioral responses to environmental conditions. The choice of a conformation for any set of environmental conditions will depend upon the robot's ability to assess those conditions, and recall previous conformations that accomplished a task objective and met the prevailing optimization criteria.

The computational resources provide the substrate for connectivity matrices between sensors and effectors. These matrices are composed of fixed and modifiable (plastic) elements, and are graphically shown in Figure 3. The arrows of Figure 3 indicate the direction of information flow. The control laws are embedded in the two boxes labeled "fixed connections" and "plastic connections". The fixed connections are established primarily by design, while the plastic connections are established primarily through the vehicle's experience in operation, though based upon pre-defined mechanisms. Feedback is indicated in the horizontal arrows between the boxes of connections, and in the line through the environment that provides information on the physical consequences of the robot's behavior.

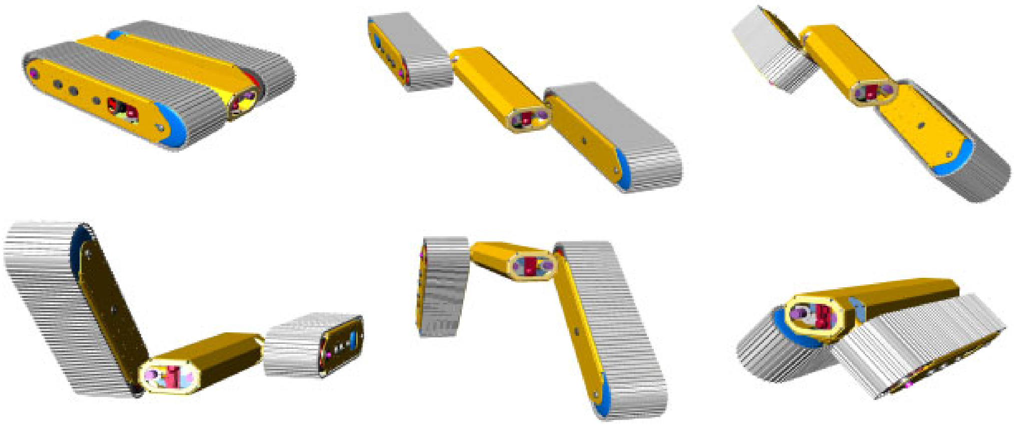

Figure 2. Various possible robot conformations.

Sensors of the internal environment include nine accelerometers (three for each of two pods, and three for the core); two angular rate-sensing gyroscopes sensitive to motion on the Z and X axes of the core; four track rotation sensors (two per pod), twenty-six IR range sensors serving as virtual whiskers sensitive at the collision limit of the vehicle (twelve located in an array on the outboard sides of each of the two pods, and two located at either end of the core faceplate;. six battery voltage sensors (two in each compartment); six battery current sensors; and six motor current sensors (one for each degree of freedom). We use a vector of features, derived from the nine accelerometers, to define the conformation of the vehicle (**C**) with respect to gravity, and as a consequence of their connectivity - to each other. The robot uses elements of this vector to make behavioral decisions.

To monitor conditions in external environments the vehicle uses the twenty-six IR range sensors looking out to 30 cm; four color video cameras, four acoustic microphones, and two RF transceivers[3].

2.1 Basic Reactive Patterns (BRP)

We provided for five basic reactive patterns that serve the survival needs of the robot and continuously modulate all overt behavior. The five BRP are continuously available to participate in the control of all overt behavior. The five BRP so far implemented are activity, balance, collision management, track contact, and energy management. The algorithms that manage the different BRP are located among the boxes labeled *fixed connections* and *plastic connections* in Figure 3.

The objective of the Activity basic reactive pattern (BRP-A) is to prevent the robot from either moving too slowly or moving too rapidly. The robot's actuators, sensors, power systems, and structural elements are all designed to function optimally within certain activity bounds. We considered these various design requirements and defined a particular scalar value with which the system will normally operate and to which it will attempt to return after disturbances. For example, the accumulation of too much rate gyro input, indicating consistently rough terrain, will slow the activity of the robot. The robot then slowly builds up through an adaptive function a counter potential to restore the original optimal performance level. Should the terrain conditions suddenly improve, the adapted performance level is higher than optimal, but the same adaptive process subsequently returns it to normal. This internal drive mechanism, sensitive to internal and

[3] The sensor side of the RF transceiver is the receiver that accepts (senses) communications from the operator control unit.

external conditions, contributes to an apparent spontaneity that permits trial and error leaning and the exercise of learned behavioral patterns.

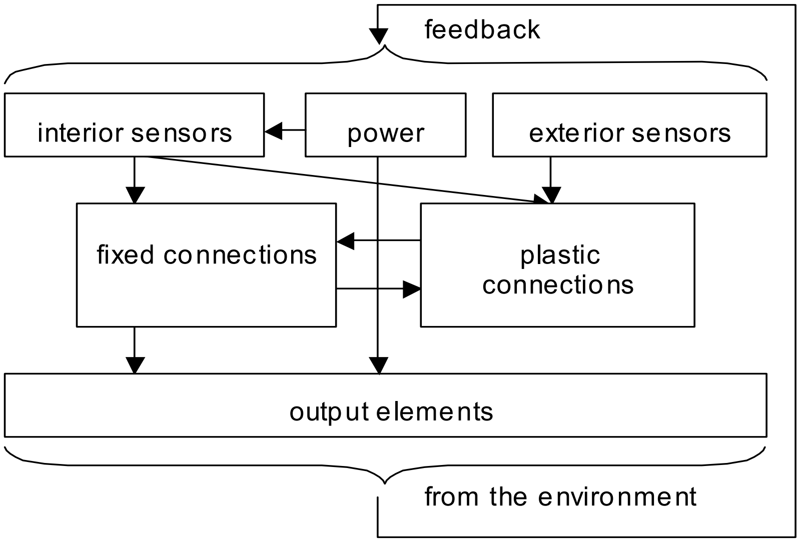

Figure 3. Schematic Architecture of the Novel UGV Control System

For the Balance basic reactive pattern (BRP-B), we have implemented a very simple algorithm that permits the robot to maintain a horizontal orientation of the core. This serves both to prevent falls and to improve operator situation awareness through the returned images from the core's video cameras. The principal feedback for this balance control comes from the core rate gyro sensors. Depending on the current conformation of the robot, motor corrections to restore balance are sent to either the pod rotation motors or to the pod camber motors. The more that the pods are oriented perpendicular to the core, the more the camber motors participate in the restoration of the core's horizontal orientation.

The Collision Management basic reactive pattern (BRP-C) extracts the robot from entanglements with objects in the external environment and protects the exposed video cameras. If the core IR range sensors fire in the absence of the pod IR range sensor activity, the BRP-C backs the vehicle away from the obstacle, using a transient inversion of the preceding motor commands. To prevent the robot from getting stuck in an infinite loop, a random potential is imposed on the subsequent forward command. The forward-looking pod IR range sensors also participate in collision avoidance by deflecting the vehicle from asymmetric obstacles. When all three forward looking IR range sensors detect an obstacle, no avoidance response is triggered, rather, the vehicle attempts to scale it by evoking the scale fixed action pattern (described below). Negative obstacles are avoided by the robot's unwillingness to go where its downward-looking IR range sensors indicate voids.

The Track Contact basic reactive pattern (BRP-D) optimizes contact with objects in the environment. A preference for objects located in the direction of gravity is built in. The BRP-D presses the track upon an object that is perceived by the IR whisker sensors to lie within reach. Pod rotation and camber are primarily employed to achieve track contact. If no contact is made, the BRP-D causes the pod to randomly explore its immediate environment in search of a contact point.

The objective of the Energy Management basic reactive pattern (BRP-E) is to acquire and conserve energy. The sensors for this BRP monitor battery voltages and current draw, and motor currents. The homeostatic tolerance for energy level is quite broad, and describes a Sigmoid similar to that for Inhibition in Figure 4. Energy acquisition behaviors (yet to be implemented) need be triggered only when energy reserves are quite low. In general, the detection of low battery charge should interrupt most on-going behavior, and trigger a recharge-specific behavior[4]. In the natural environment, with a

[4] In many interior robotic systems, an example of a recharge-specific behavior is for the robot to home on its charger and plug itself in. This would be a little more difficult to accomplish for an exterior robot operating in a complex natural environment, but the concept is the same.

limited or non-existent repertoire of navigation behaviors, the energy-limited robot may best stop all random motor activity and broadcast a call for help. With regard to energy conservation, maneuvers that draw excessive current relative to progress need to be interrupted. BRP-E will be fundamental to the learning of efficient behavioral strategies.

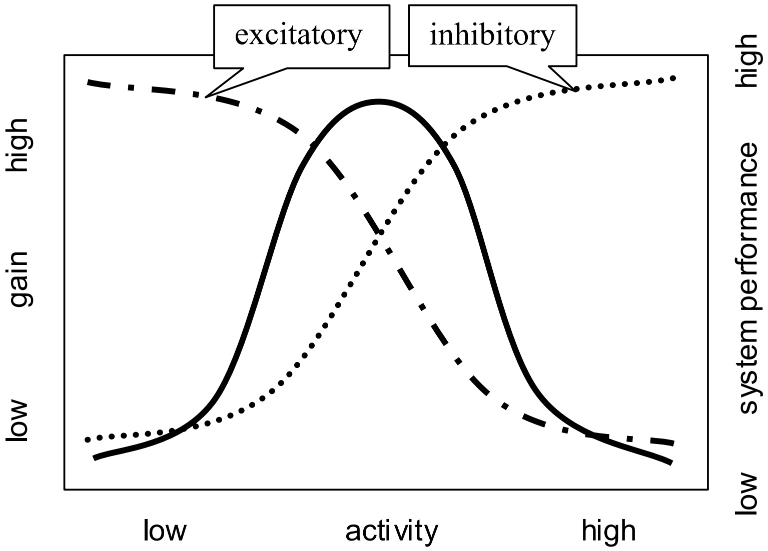

Figure 4. Relationship between preferred activity levels and performance.

Homeostasis, another term borrowed from biology, is used in this context as the aggregate of the conditions that are maintained by the basic reactive patterns above. We have quantified activity, the inverse of balance errors, the inverse of core collisions, pod contact, and energy reserve and consumption. As the BRP fail to maintain these conditions, homeostasis is upset. If you could ask the robot how it felt about the upset of its homeostasis, it might reply that it felt dysphoric. The restoration of these fundamental contributors to homeostasis will play a central role in learning and adaptation.

2.2 Fixed Action Patterns (FAP)

The robot has a repertoire of several different behaviors composed of sets of coordinated motor commands to control its six degrees of freedom during translation. Borrowing from the Neuroethology community, we call those behaviors *Fixed Action Patterns*. To date we have programmed five elementary fixed action patterns: running, scaling, undulating, walking, and yawing. The algorithms that manage the different FAP are located in the box labeled *fixed connections* in Figure 3. The Fixed Action Patterns do not necessarily depend upon any particular environmental conditions, but are invoked by triggers related to the activity in the accelerometers, rate gyros, and IR range sensors. Once initiated, FAP behavior is regulated by the changes it produces in both internal and external environments, and by the basic reactive patterns.

The Running fixed action pattern (FAP-R) permits the robot to run in a particular direction on a smooth planar surface. This FAP prefers the conformation shown in Figure 1. Sensor conditions that would favor this FAP are significant pod track contact, the absence of forward IR whisker contact, and the absence of a significant slope. To achieve this conformation, the robot rotates its pods against gravity to reduce the difference between the pod and core accelerometer values. The robot then moves forward under its own volition at a drive rate determined by its activity basic reactive pattern.

The Scaling fixed action pattern (FAP-S) permits the robot to climb a non-vertical obstacle. The robot initiates the FAP-S by encountering an obstacle with its IR whisker sensors in the forward direction of motion. The robot then rotates its two pods outward from the normal closed conformation until contact is reestablished on the pod surfaces. During rotation, the forward tracked pod normally make contact with the obstacle before the rearward pod again made contact with the ground plane, and the robot pulls itself up on the obstacle using a combination of its track tread rotations and forward track pod rotation. If the obstacle is short, the rotation could continue and the robot would pull itself over the obstacle. The BRP-B, BRP-C, and BRP-D cooperate to shape the behavior of the robot when ascending and descending stairs.

The Undulating fixed action pattern (FAP-U) permits the robot to elevate its core without moving forward. One trigger for this FAP-U is the detection of low battery capacity. (An elevated core might make the robot easier to find.) Other triggers include loss of RF signal, and loss of IR range visibility from both ends of the core. Elevation of the core could improve radio communications, and it could give the robot's video cameras a better perspective above ground rubble. Accelerometers provide the primary sensor control input during the execution of this FAP. Undulation begins from the normal closed position by rotating both pods outward, and proceeds until the core ascends to its apogee and begins again to descend. The undulation may be halted at this point whereupon the core would be at its most elevated position with respect to the ground plane. The robot is unstable in this position, but the BRP-B attempts continuously to maintain balance.

The Walking fixed action pattern (FAP-W) permits the robot to walk consistently in a particular direction on a variegated planar surface. The trigger for the FAP-W is a high level of drive and a low level of forward momentum. In this pattern, the pods rotate in the same direction, but 180 degrees out of phase, undulating the core up and down over the variegated surface. Turning on such a surface is accomplished by activating the tracks in addition to the pod rotations, by differentially rotating the pods, and by changing the camber angle of the pods. As in FAP-U, the BRP-B is critical to stability.

The Yawing fixed action pattern (FAP-Y) permits the robot to squeeze through a narrow passageway. The trigger for this maneuver is activation of the forward outboard pod whiskers and the absence of activation of the core IR range sensors. That pattern of activation indicates a gap through which the robot could attempt to squeeze. The minimum gap width that the present NGV can now negotiate is approximately eight inches. This pattern begins by the NUGV backing up and extending the pods outward as in FAP-U, however, at the point where the pods are horizontal with the core, a camber command is triggered that draws both pods in (down with respect to gravity). This maneuver forces the pods to rest on the outboard edges of their track treads. Then, alternately oscillating the ends of the pods while moving forward causes the vehicle to yaw back and forth. When the rate of yaw is correct, the vehicle should pass through an orifice of dimension down to the minimum.

2.3 Beyond the Fixed Action Patterns

We should ask at this point in our discussion of just of what is the robot capable? Given only the five Basic Reactive Patterns and the five Fixed Action Patterns so far described, we expect that the robot could self-initiate activity as its motivation for activity would initially be quite strong. We should also expect that the first FAP to be assumed would be the Run. Other FAPs may follow as conditions warrant. But Run to where? An agent with very poor external sensor capabilities may best move randomly through the environment, bouncing off this or that obstacle. Only its Basic Reactive Patterns would keep it out of trouble. Eventually though, our robot would run out of energy. The high probability for this catastrophic event is due to our design omission that does not provide the robot an opportunity to acquire energy during any FAP on its own.

The robot, as we have so far described it, is subject entirely to the fluctuations in its environments. One mechanism that nature has successfully employed to reduce this environmental subjugation, is to employ distance sensors and associate subtle changes in the external environment with significant consequential changes in the internal environment. Upon detection of those subtle changes in environmental cues, the agent can invoke a reactive process that either avoids or approaches the environmental cue. Those cues that are associated with events that restore or maintain homeostasis are fortunate for the agent. Those cues associated with events that do not must be avoided, otherwise those events will tend to terminate or exterminate the agent. Therefore, we must provide sensors of the external environment that will detect with sufficient sensitivity the subtle changes (the cues) that will predict significant change to the robot's internal environment, and we must provide a mechanism by which the robot can determine the most appropriate way to respond to those external events.

The basic purpose of these external sensors is prediction. To improve upon its homeostatic mechanisms, the robot may use its external sensors to predict the different conditions that it will encounter during its movements. We have noted that the robot's movement through the external environment engenders certain risks. Such risks are primarily related to collisions, and to the loss of contact with leverageable surfaces (e.g. falls). The external sensor information then should presage those hazards. Also, the movement of the robot may increase its likelihood of being recharged. The external sensors should detect the critical environment features that are associated with an energy source[5]. Similarly, movement itself is a homeostatic motivator, thus the external sensors should provide information that will indicate a traversable pathway (that is, one that does not impede movement).

The robot has little control over its external environment, yet its movement within that environment can change the impact that the environment might have upon it. For example, the external sensors might detect a looming object and the robot could predict a possible collision. The robot could move out of the way using similar behavioral strategies to those that it would employ had the collision been a result of its own motion through a static environment. Its avoidance of the looming object might preserve its own physical integrity, but have no effect upon the trajectory of the looming object.

Earlier in our discussion of the Fixed Action Patterns we indicated how the different patterns could be invoked by activity in the interoceptors. Ideally, the exteroceptors will provide predictive information that can be used to invoke the transformations among the Fixed Action Patterns in advance of the interoceptor triggers. In both cases, the changes in behavioral patterns should be appropriate for the conditions in the external environment, but in the second case, the robot could anticipate changes in the external environment and prepare for them. This could reduce errors and increase the speed of activity.

3. LEARNING

It is axiomatic that the measure of success for learning (long-term adaptation) is the restoration or maintenance of homeostasis. Learned behaviors are appropriate when they promote the welfare or survival of the agent, which are possible only under homeostasis. For our agent, the Novel UGV, survival may be determined by the availability of energy, by the continued operation of its hardware and software, and by its utility to the human operators. When utility disappears, the agent is subject to the trash heap. When energy dissipates, or when functionalities of hardware or of software cease, the same trash heap awaits. The learning objectives then, from the perspective of the agent, should be to maintain its energy reserves, keep itself together and functional, and meet the needs of its user. The reader may note that this last objective is something new compared to the five basic motivators discussed earlier. What will make this new objective possible is learning and long-term memory[6].

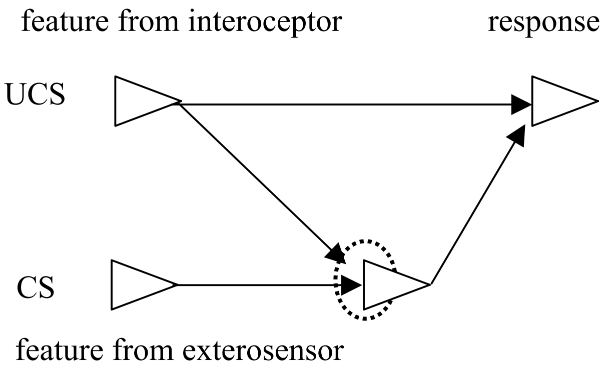

Figure 5. Simplified Classical Conditioning Paradigm

[5] For example, as the human operator is most likely to be associated with energy recovery, the robot could associate the features indicative of the presence and location of a human operator with its energy motivation, and orient to those features when energy reserves were low.

[6] The learning algorithms outlined here are not yet implemented for this application. The reader is invited to visit http://www.spawar.navy.mil/robots/pubs/PubsIdx.html for updated reports.

3.1 Classical Conditioning

During performance of a Fixed Action Pattern, the Basic Reactive Patterns will modulate the motor commands according to rules implemented in the Fixed Connection Matrix. These rules are analogous to the unconditioned stimulus-unconditioned response pairings of classical or Pavlovian conditioning. When the robot is able to perceive features of the external environment through its distance sensors, this information becomes available for association with the unconditioned response. During movement, the core accelerometers, pod pressure sensors, and faceplate pressure sensors provide the major unconditioned stimuli to support (facilitate) a conditioned response of features from the distance sensors. After conditioning (the repeated co-occurrence of the internal and external events), the features from the distance sensors invoke a response similar to the unconditioned response but in absence of the event that originally produced it. The classical conditioning paradigm is diagrammed in Figure 5. In the sequence of events during conditioning, the external event usually precedes the internal event (a likely happening because the external sensor is a distance sensor), but the record of the occurrence of the external event persists if not the event itself. When the internal event occurs it evokes a predictable response to restore homeostasis. The persistent trace of the external event becomes associated with the response evoked by the internal event according to the mechanism of *activity dependent facilitation*.

3.2 Activity Dependent Facilitation

A general learning law, known as activity dependent facilitation (Kandel and Hawkins, 1992)[7], approximates classical conditioning and is useful in determining the contributions of a particular input through its modifiable connection to an integrating element preceding an output decision. The law is as follows:

$$\Delta w = G * ((z/ e) * w) * (a *(S - m) *(C - w) - m *(w - c))$$

where *w* is its current connection strength, *z* is the activity on the input element in question, *e* is the sum of inputs from all cooperating elements to the integrating element (prior to their filter by the *w* vector), *a* is the total activation of the integrating element (equivalent to the product of the *e* vector and the *w* vector), *S* is a constant representing the maximum permissible sum of weights connecting to any one element, *m* is the current sum of weights making contact with the integrating element, *C* is a constant representing the maximum permissible weight, *c* is a lower limit on the weight to prevent it from disappearing completely if rarely used, *G* is a constant = 1 / (S*C). When both *z* and *a* are present, *w* is increased, but when *z* appears alone, *w* is decreased.

The influence of the unconditioned stimulus in the above learning law is incorporated into the sum of inputs on the integrating element. The connection weights for the UCS are strong, not modifiable over the short term, and reliably invoke an output decision in the absence of any other cooperating inputs.

In classical conditioning, novel information from the external environment acquires the strength to evoke responses that already exist in the agent's repertoire and are appropriate for the general conditions that the novel information predicts. Additional information on the application of this learning model is available in Blackburn and Nguyen (1994)[8].

3.3 Operant Conditioning

The post-hoc appropriateness of any particular behavior is determined by factors that change the sensor values, and, in effect, indicate the change in probability of catastrophe. Our second axiom is that the Basic Reactive Patterns of behavior operate to reduce the probability of catastrophe. Thus, the Basic Reactive Patterns show the Adaptive Behavioral Controller how to operate in order to restore homeostasis. That is, when a behavioral action initiated by some command from the Adaptive Behavioral Controller results in an internal sensor reading that indicates that a) activity is restored to its midrange, b) balance is restored, c) collisions are avoided, d) track contact is improved, and/or e) energy reserves and/or energy conservation are improved, we can be assured that the probability of a catastrophe has been reduced. These successful behaviors under the given environmental conditions, should be remembered so that they can be

[7] E.R. Kandel and R.D. Hawkins (1992) The biological basis of learning and individuality. *Scientific American*, 267, 78-86.

[8] M.R. Blackburn and H.G. Nguyen, H.G. Learning in robot directed reaching: A comparison of methods. *Proceedings of the 1994 Image Understanding Workshop*, Monterey, CA. Nov 13-16, 1994, 781-788.

repeated whenever the appropriate conditions reappear. Similarly, when a behavioral action results in too much or too little activity, loss of balance, collisions, loss of track contact, and depleted energy, that behavior should also be remembered and inhibited whenever those prevailing environmental conditions reappear[9].

When a specific motivator is out of homeostatic bounds, the previously associated behaviors should be primed for action. An efficient way to accomplish this priming is through the association of the interceptor features with features from the exteroceptors. The biasing of the exteroceptor features would in turn bias specific behaviors when the environment contained stimuli characterized by those features.

Our third axiom is that all acquired behavior for our robot will be expressed through the modulation of the Fixed Action Patterns using pathways in parallel with the five Basic Response Patterns that also modulate the FAP.

The locus of learning in our control architecture of Figure 3 will be the box labeled *plastic connections*. The reader will notice that this box receives input from the internal sensors, the external sensors, and the box containing fixed connections. Recall that the FAP are generally modifications of the FAP-R that is executed while the robot is in its normal closed conformation. The robot expands from this conformation to adapt primarily to information from its immediate neighborhood sensed by the IR and Whisker sensors. Recall also that the BRP generally motivate and modify the FAP based upon information from the sensors that are monitoring the internal environment. Thus, through our external influences on the stimuli that control the BRP, we can intervene and modulate any motor command associated with any FAP during performance.

Evidence that learning has occurred will be a modification of a FAP that is not immediately predicted by a complete knowledge of the internal and external environment, for learning will have permitted the robot to predict and precede an environmental event with a unique behavior.

For those readers familiar with the biological Learning Literature, we will implement here analogues of instrumental (or operant) conditioning, also known as *reinforcement learning*. Like classical conditioning, reinforcement learning requires the agent's perception of environmental information. In addition, operant learning requires an action on the part of the agent separate from the unconditioned response, and it requires some perceivable consequences of that action. The agent can use any of the available sensor information for the assessment of the environment and for the assessment of its behavioral consequences.

The process and rules of reinforcement learning that we can implement are as follows:

- Assess the internal environment (I)
- Assess the external environment (E)
- Perform an action (A)
- Reassess the internal environment, and determine if homeostasis (H) is improved.
- If H is improved, then associate factors I, E, and A, such that if I and E, then facilitate A.
- If H is worsened, then associate factors I, E, and A, such that if I and E, then inhibit A.

The above rule suggests that our controller have a special circuit that can inhibit or veto a particular action. This circuit may participate in the association rule above whenever homeostasis is disturbed by a behavior. The rules for operant conditioning are graphically represented in Figure 6. In Figure 6, arrowheads indicate direction of information flow. The line terminating in a dot represents inhibition. The dotted circles represent locations of activity dependent facilitation or inhibition.

The director, serving in this case as the supervisor of learning, need not go to great lengths to manipulate the environment in order that specific changes in homeostasis accompany particular actions under those conditions. This is because the learning algorithm above guarantees that the probability of occurrence of a particular action in the future will depend upon the prevalence of those specific internal as well as external environmental conditions. In the future, when the director may wish to see that particular action in response to particular external conditions, the internal conditions

[9] The exclusive-or problem that is solvable by a three-layer Perceptron is an example of a two choice paradigm where one choice must be inhibited in favor of the alternative under the co-occurrence of two otherwise permissible stimuli.

may not be present with sufficient intensity to drive the action above behavioral thresholds or above competing behaviors. Thus the director should generally not mess with the internal conditions of the robot.

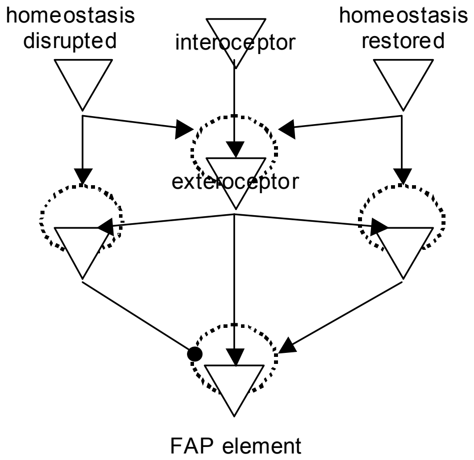

Figure 6. Simplified Operant Conditioning Paradigm.

In operant conditioning, novel information from the external environment acquires the strength to evoke responses that already exist in the agent's repertoire, but that were previously unrelated to any intrinsic motivators.

4. CONCLUDING COMMENTS

The *reinforcement learning* algorithm above requires that an action take place before the test of homeostasis. Before learning, the only behaviors of which the robot is capable are the fixed action patterns. Thus the robot will be performing a fixed action pattern when learning initially takes place. Learning will modify the particular FAP and invoke that modified FAP pattern in the future whenever the associated internal and external environmental conditions are present. When the environment is novel, the agent will default to previously learned behaviors or to the original FAP, depending upon the degree of novelty and motivation.

After some modifications of the five FAP, the repertoire may be expanded with new sequences of actions by building upon the previous action patterns that are invoked by the prevailing environmental conditions. This process is known as behavioral shaping and permits learning to progress without destroying previously learned patterns. In this way the repertoire could become quite complex, depending upon the agent's ability to discriminate the necessary behavior specific features from the external environment, and upon its ability to respond differentially to those features.

The five FAP exercise most of the mobility degrees of freedom of the robot in coordinated patterns that accomplish mobility under a variety of external conditions. The BRP provide transitory modifications to the coordinated FAP to meet certain exigencies and promote homeostasis. The external sensors can extend through classical conditioning the range of events through which the BRP are active. With any given external environment, and sufficient range of sensitivity in the external sensors, the modifications to the FAP, and even the switching among them, can create the impression of the invention of novel capabilities, when in fact, only old capabilities are being rearranged.

Operant conditioning provides for additional control flexibility that increases the potential for behavioral unpredictability given any current set of environmental conditions. Previous associations, established through operant conditioning and preserved in memory, will cooperate to produce current behaviors. The predictions and consequent behaviors that result may not always be confirmed by the evolving environmental conditions. That is, the robot may make mistakes. However, it may also make new and successful responses, demonstrating apparent creativity, and as each successful response will be in turn reinforced and saved in memory for the next opportunity for expression, demonstrate intelligence.